情報幾何学と平均場近似(多体問題としての情報処理-統計力学と情報科学の接点-,研究会報告)

田中, 利幸

情報幾何学と平均場近似

東京都立大学大学院工学研究科　　田中 利幸 [1]

1　はじめに

学習，パターン分類，画像処理などの情報処理の問題はしばしば，統計的推測の問題として定式化でき，統計学のさまざまな手法が利用される．問題の規模が大きくなると，「大自由度の統計学」としての統計力学の枠組みが自然に適用できる場合がある．統計力学の分野で開発されてきたさまざまな手法を駆使して，情報処理の問題に対して多くの研究が行われている．例えば，学習，符号，画像復元などの問題に対してレプリカ法などによる解析がなされており，それぞれの問題の性質について多くのことが明らかにされている．本稿で扱う平均場近似もまた，統計力学の手法が情報処理の問題に応用されている一例である．平均場近似は，問題の性質を解析するために使われることもあるけれども，むしろ問題を実際的に解くための効果的な近似アルゴリズムを与える点において重要である．

情報処理の分野で使われている平均場近似は，ほとんどがナイーブな平均場近似である．また，その理論的な側面に関しては，情報処理の立場からはほとんど注意が向けられてこなかった．他方，統計力学の分野では，Thouless-Anderson-Palmer (TAP) の方法 [1] をはじめとしてより進んだ平均場近似に関する理論が開発されている．けれども，それらの内容の統計学的な意味を正しく理解することは容易ではなく，そのことが，これらの進んだ平均場近似が情報処理の分野において広く活用されることを妨げているように思われる．著者は，Plefka 展開 [2] にもとづく平均場近似の定式化に対して，情報幾何学 [3, 4] にもとづく解釈を与えた [5, 6]．情報幾何学は，統計モデルの族を多様体としてとらえ，その幾何学的構造を議論することでモデルの統計的性質を明らかにするための理論的基礎を提供している．平均場近似についても，情報幾何学にもとづいてその統計学的な意味がより明確にされていくのではないかと期待できる．本稿では，平均場近似に対して情報幾何学にもとづく統計学的な解釈について現在までに得られている知見，およびそこから派生してくるいくつかの問題について述べる．

[1]E-mail: tanaka@eei.metro-u.ac.jp

2 問題

本稿で取り扱う問題を述べる．変数 $\boldsymbol{s} = (s_1, \ldots, s_N)$ に対してハミルトニアンを

$$H(\boldsymbol{s}) = -\sum_i h_i s_i - \sum_{\langle ij \rangle} w_{ij} s_i s_j \tag{1}$$

と定義する． $H(\boldsymbol{s})$ から得られるボルツマン・ギブス分布 $q(\boldsymbol{s})$ は，統計学的にいえば，$\boldsymbol{h} = (h_i), \boldsymbol{w} = (w_{ij})$ をパラメータとする指数分布族 $\mathcal{B}$ を定義する．平均場近似の対象となる問題は，パラメータ $\boldsymbol{h}, \boldsymbol{w}$ から期待値 $\boldsymbol{m} \equiv \langle \boldsymbol{s} \rangle_q$ を求める問題である． $\boldsymbol{s} \in \{-1, 1\}^N$ などの場合には，直接の計算によって期待値を求めるならば変数の数 N の指数オーダーの計算量が必要となり，実用的には何らかの近似手法が必要となる．平均場近似はこの問題に対する近似解法を与える．

3 平均場近似

平均場近似を得る方法はたくさんあるが，本稿ではふたつの方法を取り上げる．ひとつは「変分原理」にもとづくもの [7] であり，もうひとつは「摂動展開」にもとづくもの [2] である．

変分原理にもとづく導出では，相互作用の係数 w_{ij} がすべて 0 であるような，$\{\theta_i\}$ をパラメータとするハミルトニアン

$$H_0(\boldsymbol{s}) = -\sum_i \theta_i s_i \tag{2}$$

を考え， $H_0(\boldsymbol{s})$ から得られるボルツマン・ギブス分布の族 $\mathcal{F}_0$ によって q を近似することを考える． Kullback ダイバージェンス $D(p\|q) = \langle \log(p/q) \rangle_p$ を最小とする $p \in \mathcal{F}_0$ を求めれば，それが q に対する「最良」の近似だということになる． $\langle \boldsymbol{s} \rangle_p$ をもって $\boldsymbol{m}$ の近似とする． $D(p\|q)$ の停留条件として平均場方程式が得られる．これは，いわゆるナイーブ平均場近似にあたる．

変分原理にもとづく平均場近似の導出は，現在では情報処理の文脈でもよく理解されていると言ってよい．グラフィカルモデルをはじめとする多くの統計的モデルに対して，変分原理にもとづいて平均場近似が定式化され，適用されている [7].

摂動展開にもとづく導出では， q に関するギブス自由エネルギー G を， $\boldsymbol{m}$ を固定して相互作用の係数 $\boldsymbol{w}$ に関してテイラー展開する[2]．このテイラー展開は， Plefka 展開 [2] とよばれる．テイラー級数を 1 次までで打ち切ればナイーブ平均場近似が得られる．より高次の項まで考慮することによって，高次の平均場近似が得られる [9]． TAP の方法 [1] は，熱力学的極限において主要な寄与をする項をすべて残して議論するというアプローチである．こうして得られた近似的ギブス自由エネルギーについて，停留条件を考えることによって平均場方程式が得られる．

[2] $1/N$ に関する展開による導出 [8] などもあるが，ここでは触れない．

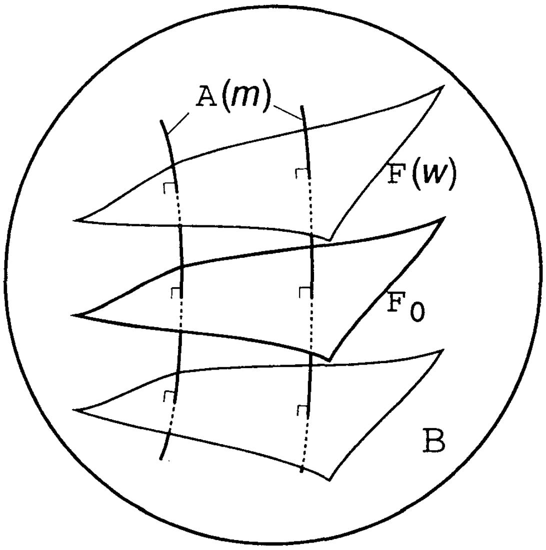


図 1: $\mathcal{B}$ の直交双対な葉層構造 $(\mathcal{F}, \mathcal{A})$ の模式図

摂動展開にもとづく平均場近似の導出は，情報処理の文脈ではそれほど浸透しているとはいえない．Opper and Winther[10] は，ガウシアンプロセスにおけるベイズ識別の問題に対して，Plefka 展開にもとづく平均場近似を適用している．

4 情報幾何学による解釈

変分原理および摂動展開の，どちらの導出に対しても，情報幾何学 [3, 4] にもとづく幾何学的な解釈を与えることが可能である．

情報幾何学の立場では，$\mathcal{B}$ は「統計多様体」とよばれる微分可能多様体とみなされる．$(\boldsymbol{h}, \boldsymbol{w})$ は統計多様体 $\mathcal{B}$ のひとつの座標系であり，正準パラメータとよばれる．$\mathcal{F}_0$ は $\boldsymbol{w} = \mathbf{0}$ で特徴づけられる $\mathcal{B}$ の部分多様体である．変分原理における，$D(p\|q)$ を最小とする $p \in \mathcal{F}_0$ を求める問題は，幾何学的には q から $\mathcal{F}_0$ へ，ある意味で「垂線」を下ろすことに相当する [11]．垂線の足が p を与える．$\mathcal{F}_0$ は $\mathcal{B}$ の部分多様体として一般には平坦でなく曲がっているので，垂線の足は一般には唯一であるとは限らない．

摂動展開に対する情報幾何学的解釈は，$\boldsymbol{w}$ および $\boldsymbol{m}$ による $\mathcal{B}$ の直交双対な葉層構造 $(\mathcal{F}, \mathcal{A})$ (図 1) を導入することによって与えられる [5, 6]．q と同じパラメータ $\boldsymbol{w}$ をもつ分布 ($\boldsymbol{h}$ は異なってもよい) の集合を $\mathcal{F}(\boldsymbol{w})$ とすると，これは $\mathcal{B}$ の部分多様体である．$D(p\|q)$ を最小にする $p \in \mathcal{F}(\boldsymbol{w})$ を求める問題を考える．この問題は自明な解 $p = q$ をもつが，この事実にはしばらく目をつぶることにする．p を指定するためには独立なパラメータとして正準パラメータ $(\boldsymbol{h}, \boldsymbol{w})$ をとるのが自然ではあるが，摂動展開の情報幾何を考えるためには，p を指定する独立なパラメータとして直交双対な葉層構造 $(\mathcal{F}, \mathcal{A})$ と対

応した座標系 $(\boldsymbol{m}, \boldsymbol{w})$ をとる．$\boldsymbol{m} \equiv \langle \boldsymbol{s} \rangle_p$ である．$(\boldsymbol{m}, \boldsymbol{w})$ は混合座標系とよばれており，(少なくとも有限の N に対しては) 正準パラメータ $(\boldsymbol{h}, \boldsymbol{w})$ と同様に $\mathcal{B}$ の座標系として使うことができる．$D(p\|q)$ を $\boldsymbol{m}$ を固定して $\boldsymbol{w} = \boldsymbol{0}$ に関してテイラー展開する．すると，このテイラー展開が，Plefka 展開を与えていることがわかる．ナイーブ平均場近似との関係や，線形応答定理による相関の推定などについても，情報幾何学的の立場から統一的な解釈を与えることができる [5, 6].

5 Plefka 展開の収束性

Plefka 展開は $\boldsymbol{w}$ に関するべき級数展開であるから，一般のべき級数と同様に，収束域が定義される．数学的には，$\boldsymbol{w}$ が収束域に含まれている場合には Plefka 展開は意味をもち，級数を有限項で打ち切ることによって得られる近似は，数学的にみて「よい」近似を与えることが期待できる．けれども，収束域の外の $\boldsymbol{w}$ に対しては，Plefka 展開はべき級数として収束せず，平均場方程式の意味も数学的にはもはや明確であるとはいえなくなる．したがって，数学的にいえば摂動展開によって得られる平均場近似は Plefka 展開の収束域でのみ有効であるということができる．

SK モデルについては，Plefka 展開の収束域は Plefka[2] によって研究されている．$w_{ij} \sim N(0, 1/N)$ とすると，Plefka 展開が収束する条件は $N \to \infty$ で

$$\beta^{-2} > \max\{(1 - 2q_2 + q_4), 2(q_2 - q_4)\} \tag{3}$$

$$q_n \equiv \frac{1}{N} \sum_i m_i^n \tag{4}$$

で与えられる．この収束条件と，レプリカ法による解析結果とを比較するために，$h_i \equiv h$ とすると，同一視

$$q_n = \int Dz \tanh^n(\beta(q_2^{1/2} z + h)) \tag{5}$$

のもとで式 (3) の第一項は SK 解に対する de Almeida-Thouless (AT) 線 [12] に対応していることがわかる．レプリカ法による解析結果からは，第二項に対応する条件は得られない[3]．h-$1/\beta$ パラメータ平面において SK 解について調べてみると，図 2 にみられるように，$1/\beta$ が小さい部分で式 (3) の第二項が限界を与えることがあるので，SK 解に対応する解における Plefka 展開の有効性が数学的に保証される範囲は RS 領域に含まれ，それより本質的に狭いことになる[4]．

情報処理の問題への応用の観点から重要なのは N が有限の場合である．この場合にはモデルはエルゴード的であるが，$D(p\|q)$ は $\boldsymbol{w}$ の複素数値に対して特異性をもち，その特異点が Plefka 展開の収束域を制限する．この制限が実際の平均場近似の応用に対してどのように影響するかを調べることは，興味深い課題である．

[3]第二項は，$O(1)$ の個数のモードの不安定化に関する条件なので，鞍部点近似によって $O(1)$ の項を落としてしまうレプリカ法の解析結果には現れてこないものと考えられる．

[4]図 2 からもわかるように，第二項の寄与による収束域の狭まりはごくわずかである．

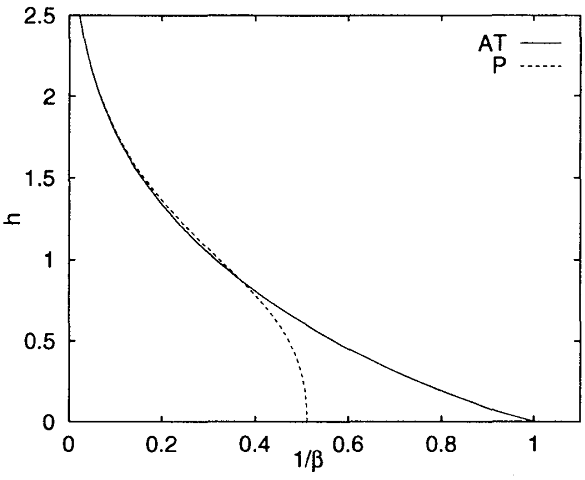


図 2: h-$1/\beta$ 平面における SK 解に対する Plefka 展開の安定性条件．実線 (AT)，波線 (P) は式 (3) 右辺のそれぞれ第一項，第二項から決まる安定限界をあらわす．

6 非エルゴード性と情報幾何学

6.1 平均場近似の統計学的解釈をめぐる問題点

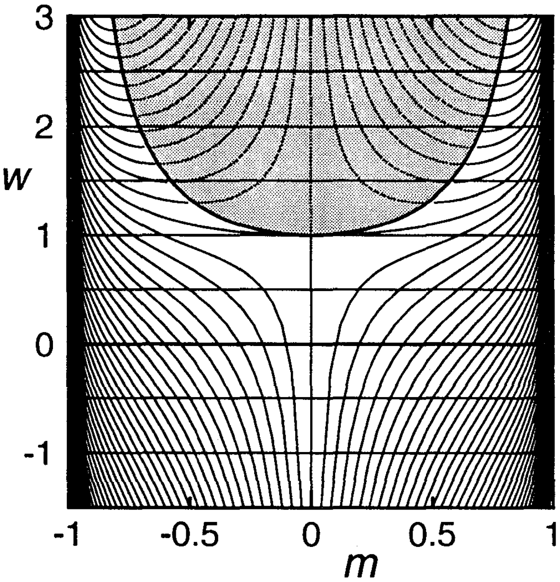
SK モデルについての研究において，平均場方程式は AT 相転移温度以下の低温では解を複数もつことがよく知られている．ところが，情報幾何学の立場からみると話は違ってくる．多様体 $\mathcal{F}(\boldsymbol{w})$ は統計多様体として平坦であることが容易にわかるので，対応する最小化問題

$$\min_{p\in\mathcal{F}(\boldsymbol{w})} D(p\|q) \tag{6}$$

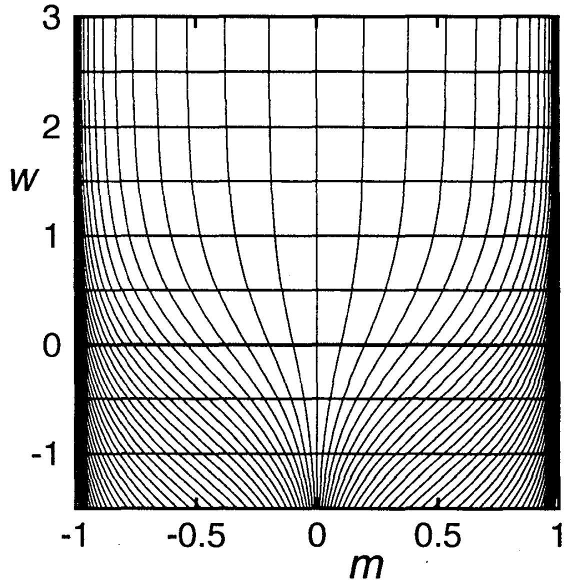
の解は自明な解 $p=q$ だけであることが言える．両者のあいだには，明らかな矛盾がある．

この矛盾から，重要な問題が提起される．統計力学の立場では，平均場方程式が多数の解をもつのは，系がエルゴード性を喪失していることと対応するとされており，ひとつのエルゴード成分に関しての期待値は，平均場方程式のあるひとつの解に対応づけられる，ということが基本的には広く信じられている．けれども，平均場方程式の解の性質は物理学においても完全に理解されているわけではない [13]．とくに，情報処理への応用を想定する立場からすれば，平均場方程式の解に対する「統計学的な」解釈を与えることはたいへん重要な問題である．

Plefka 展開による平均場近似の導出に対しては情報幾何学からの解釈が与えられたのだから，平均場方程式が複数解をもつときのそれらの解に対する統計学的な解釈についても，情報幾何学的な考察から何らかの手がかりが得られるのではないか，ということが期待される．ここでの本質的な問題は，熱力学的極限においてはじめて出現する非エルゴード性であり，そのような状況が情報幾何学においていかに表わされるかを調べる必要があ



(a) 強磁性モデル



(b) $N = 2$

図 3: 混合座標系 (m, w) における等 h 線

ると考えられる．そのような方向を狙って，非エルゴード性を示すもののなかでももっとも単純でその性質がよく理解されているモデルとして強磁性モデルをとりあげて，情報幾何学の立場からの検討を試みる．

6.2 強磁性モデル

本稿での問題の定式化に沿っていえば，強磁性モデルは，$w_{ij} \equiv w/N$, $h_i \equiv h$ であるモデルとして特徴づけられる．これは，各 N に対して，(h, w) を正準パラメータとする 2 次元の平坦な統計多様体を定めていることになる．Plefka 展開を考えると，$\boldsymbol{w}$ に関して 2 次以上の項は $N \to \infty$ の極限で無視でき，よく知られているように $N \to \infty$ の極限において厳密な平均場方程式

$$m_i \equiv m = \tanh(wm + h) \tag{7}$$

が得られる．

強磁性モデルに対応する統計多様体の構造を調べると，これがよく知られたカスプ形の特異性をもつことがわかった．図 3 (a) に，強磁性モデルに関して混合座標系 (m, w) に示した等 h 線を示す．図中灰色の領域は，Plefka 展開の収束域外であることを示している．比較のために，図 3 (b) に，$N = 2$ のモデルに関する等 h 線を示す．$N = 2$ の場合には，各 w に対して m と h とのあいだには 1 対 1 の対応があるが，$N \to \infty$ の極限では，図 4 に模式的に示したように，$w > 1$ に対してそのような 1 対 1 対応が失われており，m が $\mathcal{F}(w)$ の座標系として機能しなくなる，ということが起こっている．

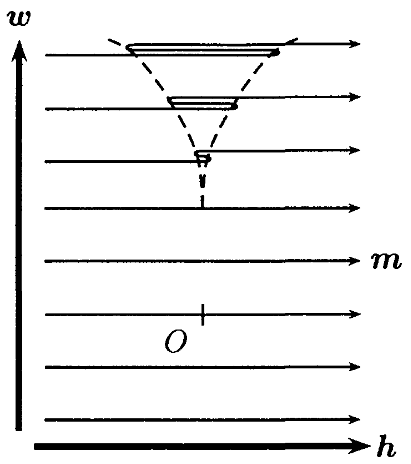


図 4: 強磁性モデルに関する統計多様体のもつカスプ特異性の模式図

以上のことは，物理的には「常磁性 - 強磁性相転移にともなう自発磁化の発生」としてよく知られていることである．けれども，情報幾何学的には「統計多様体の大域的な幾何学構造の変化」であり，その意味するところは今のところ明確であるとはいえない．統計力学では普通に現れるこの種の非エルゴード性が，情報幾何学においてどのように表現されるのか，という問題は，これまでのところまったく研究されておらず，たいへん興味深い問題である．

7 まとめ

本稿では，Plefka 展開による平均場近似の導出に対して情報幾何学の立場からの解釈を与えた著者の研究の概略を紹介し，そこから派生する諸問題について述べた．

参考文献

[1] D. J. Thouless, P. W. Anderson and R. G. Palmer, *Phil. Mag.*, **35** (1977), 593.

[2] T. Plefka, *J. Phys. A: Math. Gen.*, **15** (1982), 1971.

[3] S. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, vol. 28, Springer-Verlag (1985).

[4] 長岡，甘利，「情報幾何の方法」，岩波講座 応用数学，岩波書店

[5] T. Tanaka, in M. S. Kearns et al. (eds.), *Advances in Neural Information Processing Systems*, vol. 11, The MIT Press (1999), pp. 351–357.

[6] T. Tanaka, to appear in *Neural Computation* (1999).

[7] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, in M. I. Jordan (ed.), *Learning in Graphical Models*, Kluwer Academic Publishers, Dordrecht (1998), pp. 105–161.

[8] K. Nakanishi, *Phys. Rev. B*, **23** (1981), 3514.

[9] T. Tanaka, *Phys. Rev. E*, **58** (1998), 2302.

[10] M. Opper and O. Winther, to appear in M. S. Kearns et al. (eds.), *Advances in Neural Information Processing Systems*, vol. 11, The MIT Press (1999).

[11] T. Tanaka, *IEICE Trans. Fundamentals*, **E79-A** (1996), 709.

[12] J. R. L. de Almeida and D. J. Thouless, *J. Phys. A: Math. Gen.*, **11** (1978), 983.

[13] G. Parisi and M. Potters, *J. Phys. A: Math. Gen.*, **28** (1995), 5267.